



**HAL**  
open science

## 10 ans d'évolution des pratiques de science ouverte en Sciences Humaines et Sociales vues de l'infrastructure de recherche Huma-Num

Nicolas Larrousse, Olivier Baude

► **To cite this version:**

Nicolas Larrousse, Olivier Baude. 10 ans d'évolution des pratiques de science ouverte en Sciences Humaines et Sociales vues de l'infrastructure de recherche Huma-Num. *Culture et Recherche*, 2023, 144, pp.11. halshs-04380681

**HAL Id: halshs-04380681**

**<https://shs.hal.science/halshs-04380681>**

Submitted on 8 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC0 - Public Domain Dedication 4.0 International License

## ***10 ans d'évolution des pratiques de science ouverte en Sciences Humaines et Sociales vus de l'infrastructure de recherche Huma-Num<sup>IR\*</sup>***

### **Auteurs :**

Nicolas Larrousse : Huma-Num CNRS UAR 3598 - ORCID [0000-0002-4968-797X](https://orcid.org/0000-0002-4968-797X)

Olivier Baude : Huma-Num CNRS UAR 3598 - ORCID [0000-0002-8627-5229](https://orcid.org/0000-0002-8627-5229)

L'infrastructure Huma-Num<sup>IR\*</sup> a été créée il y a dix ans par la fusion de deux structures. L'une proposant des services technologiques, le TGE (Très Grand Équipement) Adonis, et l'autre assurant la coordination de communautés organisées en consortiums autour de la création de jeux de données numériques, l'IR (Infrastructure de Recherche) Corpus. L'infrastructure est inscrite sur la feuille de route nationale des infrastructures du MESR (Ministère de la Recherche et de l'Enseignement Supérieur). Son label d'IR\* (ex TGIR) indique qu'elle relève d'une politique nationale et d'un fléchage budgétaire du MESR.

La notion d'infrastructure en SHS était à cette époque un concept qui restait largement à imaginer. L'organisation originale de l'IR\* Huma-Num<sup>1</sup>, issue de sa double origine qui associe un dispositif technologique aux communautés scientifiques, a permis de s'appuyer sur des compétences extérieures pour guider l'évolution de l'infrastructure au plus près des besoins des communautés. A cet égard, les Consortiums-HN<sup>2</sup> (*cf illustration 1*) labellisés et soutenus par Huma-Num<sup>IR\*</sup> ont joué un rôle essentiel dans la construction de l'infrastructure.

Au commencement, le besoin principal concernant les données était surtout de ne pas les perdre, compte tenu du fait qu'elles se trouvaient stockées sur des supports fragiles comme des disques optiques, des clefs USB, des disques durs portables voire de supports plus exotiques (eg disquettes ou autres cartouches). Cependant, il est apparu rapidement que préserver simplement les données brutes n'était pas suffisant. En effet, il est inutile de conserver des données qui potentiellement ne seront plus lisibles car encodées avec des formats non pérennes, c'est-à-dire des formats qui ne sont pas documentés, qui ne sont pas outillés ou plus généralement dont les communautés utilisatrices ne sont plus actives. Les Sciences Humaines et Sociales sont constituées de différentes disciplines qui utilisent des formats très diversifiés, et l'équipe technique d'Huma-Num<sup>IR\*</sup> ne peut disposer de l'expertise sur tous ces domaines. Afin d'accompagner les choix de formats effectués, Huma-Num<sup>IR\*</sup> bénéficie des compétences développées au sein de ses Consortiums-HN, associées à celles de ses partenaires comme le CINES pour la préservation pérenne. Par exemple, le travail conjoint avec le consortium 3D a abouti à la création d'un jeu de métadonnées spécifiques adaptées aux objets 3D ainsi qu'au choix de formats de fichiers pour leur préservation sur le long terme. C'est ainsi que Huma-Num<sup>IR\*</sup> a pu mettre progressivement en œuvre une chaîne (*cf illustration 2*) de traitement qui couvre le cycle de vie des données de la recherche, de leur collecte à leur préservation.

---

<sup>1</sup> <https://www.huma-num.fr/>

<sup>2</sup> <https://www.huma-num.fr/les-consortiums-hn/>

L'un des éléments clef de ce dispositif est l'entrepôt NAKALA<sup>3</sup> (cf illustration 3) qui a été conçu par Huma-Num<sup>IR\*</sup> pour permettre de publier des données documentées, d'en assurer la lisibilité dans le temps et de les partager en vue de leur réutilisation. Ces caractéristiques ont permis de faciliter la mise en oeuvre des principes FAIR<sup>4</sup>, constituants centraux de la science ouverte. La découvrabilité des données et productions scientifiques est ensuite assurée par le service ISIDORE<sup>5</sup> (cf illustration 4) qui est un moteur de recherche associé à un assistant scientifique fondé sur les technologies du web sémantique et de l'intelligence artificielle.

L'enjeu aujourd'hui n'est plus le risque de perdre les données mais de pouvoir assister les projets de recherche dans l'amélioration de la qualité des données et des métadonnées qu'ils produisent. C'est dans cet objectif que Huma-Num<sup>IR\*</sup> a créé un pôle d'accompagnement qui permet d'orienter les choix de formats, de référentiels et d'organisation en vue d'une publication des jeux de données respectant des critères de qualité permettant d'en assurer la traçabilité, la citation et la découvrabilité afin d'en faciliter la réutilisation. De même que pour la création de la chaîne de traitement de données mentionnée auparavant, Huma-Num<sup>IR\*</sup> implique fortement les communautés dans ce processus. Cette implication se réalise notamment en s'appuyant sur le réseau de correspondants de l'infrastructure dans les Maisons des Sciences de l'Homme qui a été développé ces dernières années. Ce réseau<sup>6</sup> assure un relais naturel avec les projets de recherche soutenus par ces infrastructures régionales et complète les liens avec les communautés établis avec les consortiums de Huma-Num<sup>IR\*</sup>.

Durant cette décennie, le paysage national a lui aussi beaucoup évolué, en particulier récemment avec la création de l'initiative RDG<sup>7</sup> (Recherche Data Gouv) qui constitue un écosystème pour le partage et l'ouverture des données de recherche. Compte-tenu de son expérience, des dispositifs mis à disposition des communautés scientifiques ainsi que du maillage national constitué, Huma-Num<sup>IR\*</sup> a été labellisé "Centre de Référence Thématique<sup>8</sup>" pour les Sciences Humaines et Sociales. Il permet à Huma-Num<sup>IR\*</sup> de s'inscrire dans une initiative nationale et ainsi de favoriser la valorisation des données en Sciences Humaines et Sociales stockées dans l'entrepôt NAKALA, désigné entrepôt de confiance, qui sera moissonné par le catalogue national. De même, la démarche d'amélioration de la qualité des données entreprise par Huma-Num<sup>IR\*</sup> bénéficiera pleinement du dispositif d'accompagnement constitué par le réseau national des "Ateliers de la donnée<sup>9</sup>" mis en place dans le cadre de l'initiative RDG. La maîtrise de la chaîne de traitement des données au cœur des pratiques scientifiques actuelles permet également à Huma-Num<sup>IR\*</sup> de travailler avec différents partenaires, notamment avec la BnF (Bibliothèque Nationale de France) dans le cadre du BnF DataLab<sup>10</sup>, un service d'accès outillé aux collections de la BnF pour les chercheurs.

La maturité du dispositif élaboré par Huma-Num<sup>IR\*</sup> pour l'ouverture des données de recherche permet d'envisager de nouvelles possibilités de valorisation de ces productions essentielles à la démarche scientifique. Un exemple est l'établissement d'un lien permanent entre les publications et les données associées. Une première exploration de cette notion a été réalisée dans le cadre du projet européen

---

<sup>3</sup> <https://nakala.fr/>

<sup>4</sup> <https://www.go-fair.org/fair-principles/>

<sup>5</sup> <https://isidore.science/>

<sup>6</sup> <https://www.huma-num.fr/carte-des-relais-huma-num-dans-les-msh/>

<sup>7</sup> <https://recherche.data.gouv.fr/>

<sup>8</sup> <https://recherche.data.gouv.fr/fr/page/centres-de-reference-thematiques-expertises-par-domaine-scientifique>

<sup>9</sup> <https://recherche.data.gouv.fr/fr/page/ateliers-de-la-donnee-des-services-generalistes-sur-tout-le-territoire>

<sup>10</sup> <https://www.bnf.fr/fr/bnf-datalab>

EOSC-PILLAR<sup>11</sup>. Une preuve de concept de réalisation du lien entre données et publications a été construite entre l'entrepôt de données NAKALA et l'outil de dépôt de publication HAL-SHS<sup>12</sup> géré par le CCSD<sup>13</sup>. Cette expérimentation a permis d'identifier les prérequis nécessaires à la création de liens respectifs entre les données et les publications : la création de métadonnées croisées sur les deux plate-formes à la synchronisation nécessaire pour conserver la cohérence de ces liens ainsi que la publication standardisée de ces liens pour en assurer la découvrabilité.

Les enseignements de ce projet seront mis à profit dans le cadre du projet d'EquipeX+ COMMONS<sup>14</sup> qui associe trois infrastructures nationales autour de la création de lien entre les données et les publications : Huma-Num<sup>IR\*</sup> pour les données, OpenEdition<sup>15</sup> et METOPES<sup>16</sup> pour les publications. Ce projet qui a débuté en 2022 se déroulera sur huit ans. Il vise à développer des passerelles entre les plate-formes des différentes infrastructures pour faciliter l'établissement de liens entre les données et les publications pour les différentes catégories d'utilisateurs comme les éditeurs de revues ou les projets de recherche. Le développement de formation pour ces utilisateurs représente un volet important du projet ainsi que l'ajustement permanent des objectifs et des réalisations du projet par l'utilisation des résultats d'un observatoire des usages.

Comme on peut le constater, ces dispositifs de soutien pour la mise en œuvre de la science ouverte ne peuvent se développer qu'en associant étroitement les communautés utilisatrices. Par ailleurs, en raison des évolutions très importantes et rapides des pratiques durant ces dernières années, il est nécessaire de disposer d'une vue prospective sur les dynamiques en cours afin de pouvoir anticiper les usages futurs. Dans cet objectif, Huma-Num<sup>IR\*</sup> a créé un pôle qui permet d'alimenter la réflexion sur les nouveaux services à développer : le Huma-Num Lab<sup>17</sup>.

---

<sup>11</sup> <https://www.eosc-pillar.eu/>

<sup>12</sup> <https://shs.hal.science/>

<sup>13</sup> <https://www.ccsd.cnrs.fr/>

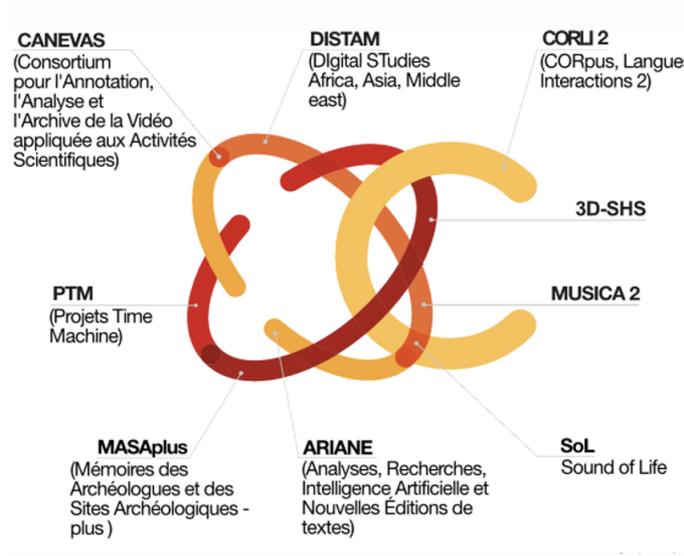
<sup>14</sup> <https://leo.hypotheses.org/17434> - <https://shs.hal.science/halshs-03881307>

<sup>15</sup> <https://www.openedition.org/>

<sup>16</sup> <http://www.metopes.fr/>

<sup>17</sup> <https://www.huma-num.fr/hnlab/>

**Illustration 1 : Les Consortiums Huma-Num**



**Illustration 2 : Le cycle de vie des données de la recherche et les services Huma-Num**



**Illustration 4 : L'entrepôt Nakala**



**Partager, publier et valoriser vos données scientifiques**

[Je dépose mes données dans NAKALA](#)

**Illustration 5 : L'assistant de recherche ISIDORE**



Votre assistant de recherche  
en Sciences Humaines et Sociales

Documents ▾ Rechercher dans les 10 398 919 documents de ISIDORE...

[Recherche avancée](#)

[i démonstration](#)